PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-097268
(43)Date of publication of application : 14.04.1998

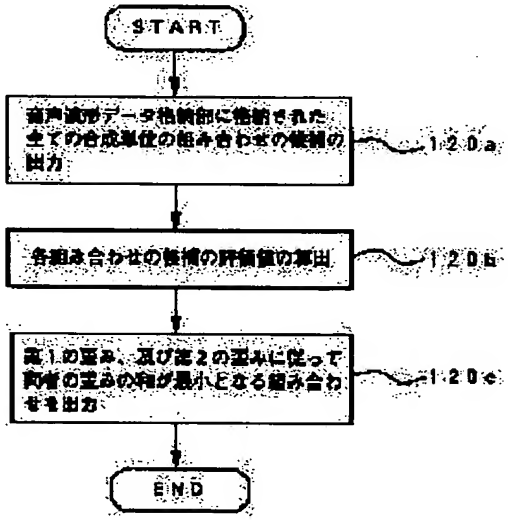
(51)Int.Cl. G10L 3/00
G10L 5/02

(21)Application number : 08-251645
(22)Date of filing : 24.09.1996

(71)Applicant : SANYO ELECTRIC CO LTD
(72)Inventor : HIRAI HIROYUKI
NISHIDA HIDEJI
ONISHI HIROKI

(54) SPEECH SYNTHESIZING DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To synthesize a voice of high quality with small auditory distortion by using not the total of distortion of connection parts in synthesis units, but the maximum value of the distortion when an optimum combination of synthesis units is selected, and selecting a combination of synthesis units which becomes small in maximum distortion value.
SOLUTION: A synthesis unit waveform selection part has a function which uses distortion (1st distortion) generated when synthesis units are connected and distortion (2nd distortion) due to a mismatch of speaking environment and evaluates whether the addition of the both is adequate for a selected combination of synthesis units. In a step 120a, candidates for combinations of all synthesis units stored in a voice waveform data storage part are extracted. In a step 12b, evaluation values of the candidates for the respective combinations are calculated. In a step 120c, the sum of the 1st distortion and 2nd distortion is regarded as an evaluation value according to the both and a combination whose value becomes minimum is outputted.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－97268

(43)公開日　平成10年(1998)４月14日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ１０Ｌ	3/00		Ｇ１０Ｌ	3/00	Ｈ
	5/02			5/02	Ｊ

審査請求　未請求　請求項の数４　ＯＬ　（全６頁）

(21)出願番号　特願平8－251645

(22)出願日　平成８年(1996)９月24日

(71)出願人　000001889
三洋電機株式会社
大阪府守口市京阪本通２丁目５番５号

(72)発明者　平井　啓之
大阪府守口市京阪本通２丁目５番５号　三洋電機株式会社内

(72)発明者　西田　秀治
大阪府守口市京阪本通２丁目５番５号　三洋電機株式会社内

(72)発明者　大西　宏樹
大阪府守口市京阪本通２丁目５番５号　三洋電機株式会社内

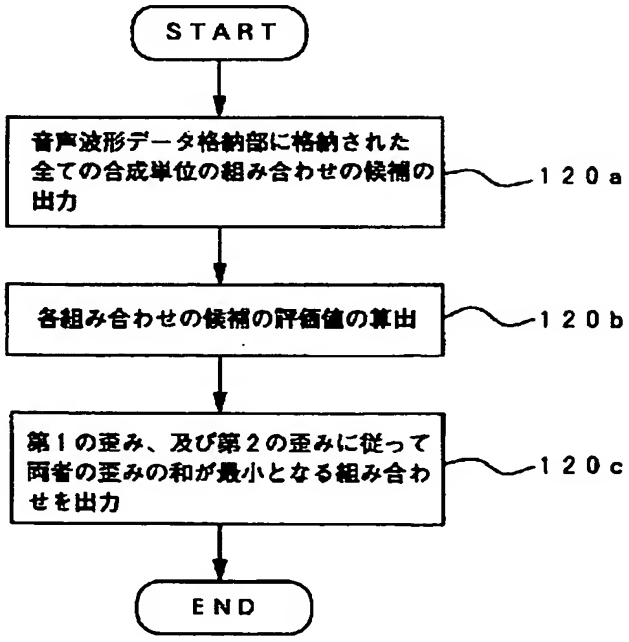
(74)代理人　弁理士　安富　耕二　（外１名）

(54)【発明の名称】　音声合成装置

(57)【要約】

【課題】　音声合成装置における波形選択に際して、動的計画法のように歪みの合計が最小になる組合せを求めたのでは、或る箇所の接続部分に歪みが集中する場合があり、その接続部分で比較的大きな雑音が発生し、その前後では歪みが小さいため、聴感的には特に大きな雑音として知覚されるという問題点がある。

【解決手段】　入力されたテキストを解析する言語処理部(10)と、予め蓄積された音声波形データ格納部(11)と、該音声波形データの中から合成単位を選択する波形接続単位選択部(12)と、選択された合成単位を接続する音声波形生成部(13)と、を備えた音声合成装置において、前記波形接続単位選択部(12)は、入力されたテキストを生成することが可能な複数の合成単位の組合せのうち、各組合せの合成単位を接続することにより生ずる歪みの最大値をその組合せの評価値とし、その評価値が最も小さい合成単位の組合せを選択することを特徴とする。



1

【特許請求の範囲】

【請求項１】　入力されたテキストを解析する言語処理部と、
音声波形データを予め蓄積している音声波形データ格納部と、
該音声波形データの中から合成単位を選択する波形接続単位選択部と、
該波形接続単位選択部にて選択された合成単位を接続する音声波形生成部と、を備えた音声合成装置において、
前記波形接続単位選択部は、入力されたテキストを生成することが可能な複数の合成単位の組合せのうち、各組合せの合成単位を接続することにより生ずる歪みの最大値をその組合せの評価値とし、その評価値が最も小さい合成単位の組合せを選択することを特徴とする音声合成装置。

【請求項２】　前記波形接続単位選択部は、２種類の歪みを選択の基準として用い、合成単位を接続することにより生じた歪みの最大値を第１の歪みとして記憶し、また各合成単位に対応する前記音声波形データの所望パラメータと該合成単位に対応する前記所望パラメータの目標値に基づく歪みの平均値を第２の歪みとして記憶し、それら両方の歪みの和をその組合せの評価値とし、その評価値が最小となる合成単位の組合せを選択することを特徴とする請求項１記載の音声合成装置。

【請求項３】　合成単位接続により生じた歪みの最大値からｎ個（ｎは整数）の歪みを記憶する記憶部を有し、入力されたテキストを生成することが可能な複数の合成単位の組合せに対し、第１の歪みのうち１番目に大きな歪みを用いて各組合せの評価値を求め、その中の最小値となる評価値から所定範囲内の評価値に属する複数の組合せを選択の候補とし、次にその複数の組合せの候補の中で第１の歪みのうち２番目に大きな歪みを用いて各組合せの評価値を求め、最も小さい評価値を与える合成単位の組み合わせによって音声合成することを特徴とする請求項２記載の音声合成装置。

【請求項４】　合成単位接続により生じた歪みの最大値からｎ個（ｎは整数）の歪みを記憶する記憶部を有し、入力されたテキストを生成することが可能な複数の合成単位の組合せに対し、第１の歪みに対して、その値の大きさの順序に応じた重み係数を掛け、それらを加算した値を各合成単位の組み合わせの評価値とし、その中の最小値となる評価値が最も小さい合成単位の組み合わせによって音声合成することを特徴とする請求項２記載の音声合成装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、入力されたテキストを解析し、音声を合成する音声合成装置において、波形接続部の歪みの最大値が小さくなる合成単位を選択して音声合成する音声合成装置に関する。

2

【０００２】

【従来の技術】従来の音声合成装置にあっては、複数の合成単位の組合せのうち最適な合成単位の組合せを選択しており、図４は、本発明を用いた音声合成装置の概略構成図を示す。

【０００３】入力されたテキストは、言語処理部（１０）で形態素解析、係り受け解析が行なわれ、音素記号、アクセント記号等に変換される。

【０００４】次に、韻律パターン生成部（１１）では、音素記号、アクセント記号列および形態素解析の結果より得られる入力テキストの品詞情報を用いて、音韻継続時間長(声の長さ)、基本周波数パターン、ピッチパターン(声の高さ)、母音中心のパワー(声の大きさ)等の推定が行われる。

【０００５】合成単位波形選択部（１２）では、音素記号列および推定された音韻継続時間長、基本周波数パターン、母音中心のパワー情報等を用いて計算された評価値に基づいて、波形辞書に蓄積されている音声波形のうち最適な合成単位の組合せが求められる。

【０００６】最後に、音声波形生成部（１３）では、選択された合成単位波形の組み合わせに従い、ピッチを変換しつつ、合成単位波形の接続を行なうことによって音声の生成を行う部分である。例えば、音声の生成に際して、PSOLA法：発表論文「Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones.」(Proc. Eurospeech' 89 (1989), Charpentier,F. and Moulines,E.)などにより実現できる。

【０００７】従来の音声合成装置では、合成単位波形選択部（１２）にて合成単位の決定に際して動的計画法（ＤＰ法）を用いて合成単位を接続しており、この接続による音声の歪みの合計を評価値として、その値が最小となる合成単位の組合せを選択していた。

【０００８】

【発明が解決しようとする課題】然し乍ら、上述の動的計画法のように音声の歪みの合計が最小になる組合せを求めたのでは、その前後の別の接続部分では歪みが小さいものの、或る箇所の接続部分に歪みが集中する場合がある。

【０００９】このような場合、或る接続部分で比較的大きな雑音が発生し、その前後では歪みが小さいため、聴感的には特に大きな雑音として知覚されるという問題点があった。

【００１０】本発明は、上述の問題点に鑑みなされたものであり、最適な合成単位の組合せを選択する基準として、歪みの合計ではなく歪みの最大値を用い、歪みの最大値が小さくなる合成単位の組合せを選択することにより音声合成を行う音声合成装置を提供する。

【００１１】

【課題を解決するための手段】本発明は、入力されたテ

キストを解析する言語処理部と、音声波形データを予め蓄積している音声波形データ格納部と、該音声波形データの中から合成単位を選択する波形接続単位選択部と、該波形接続単位選択部にて選択された合成単位を接続する音声波形生成部と、を備えた音声合成装置において、前記波形接続単位選択部は、入力されたテキストを生成することが可能な複数の合成単位の組合せのうち、各組合せの合成単位を接続することにより生ずる歪みの最大値をその組合せの評価値とし、その評価値が最も小さい合成単位の組合せを選択することを特徴とする。

【００１２】また、前記波形接続単位選択部は、２種類の歪みを選択の基準として用い、合成単位を接続することにより生じた歪みの最大値を第１の歪みとして記憶し、また各合成単位に対応する前記音声波形データの所望パラメータと該合成単位に対応する前記所望パラメータの目標値に基づく歪みの平均値を第２の歪みとして記憶し、それら両方の歪みの和をその組合せの評価値とし、その評価値が最小となる合成単位の組合せを選択することを特徴とする。

【００１３】また、合成単位接続により生じた歪みの最大値からｎ個（ｎは整数）の歪みを記憶する記憶部を有し、入力されたテキストを生成することが可能な複数の合成単位の組合せに対し、第１の歪みのうち１番目に大きな歪みを用いて各組合せの評価値を求め、その中の最小値となる評価値から所定範囲内の評価値に属する複数の組合せを選択の候補とし、次にその複数の組合せの候補の中で第１の歪みのうち２番目に大きな歪みを用いて各組合せの評価値を求め、最も小さい評価値を与える合成単位の組み合わせによって音声合成することを特徴とする。

【００１４】更に、合成単位接続により生じた歪みの最大値からｎ個（ｎは整数）の歪みを記憶する記憶部を有し、入力されたテキストを生成することが可能な複数の合成単位の組合せに対し、第１の歪みに対して、その値の大きさの順序に応じた重み係数を掛け、それらを加算した値を各合成単位の組み合わせの評価値とし、その中の最小値となる評価値が最も小さい合成単位の組み合わせによって音声合成することを特徴とする。

【００１５】

【発明の実施の形態】本発明の実施の形態を図１～図３を用いて説明する。

【００１６】本発明の音声合成装置の概略構成は図４に示した構成と基本的に同様であるが、本発明が従来の音声合成装置と異なる点は、合成単位波形選択部（１２）に代えて合成単位波形選択部（１２０）を用いたことである。

【００１７】合成単位波形選択部（１２０）は、合成単位を接続することによって生じる歪み（第１の歪み）、及び発話環境の非適合による歪み（第２の歪み）を用い、これら両者を加え合せたものが選択された合成単位の組み合わせとして適切であるか否かを評価する機能を有する。

【００１８】図１は、本発明における合成単位波形選択部（１２０）の処理の流れを示したものである。

【００１９】図１において、ステップ１２０ａでは、音声波形データ格納部に格納された全ての合成単位の組み合わせの候補を抽出する。次にステップ１２０ｂでは、各組み合わせの候補の評価値を算出する。

【００２０】ステップ１２０ｃでは、第１の歪み、及び第２の歪みに従って、それらの歪みの和を評価値とし、その値が最小となる組み合わせを出力する。

【００２１】本発明の実施の形態では、波形の接続部に歪みが集中することを抑制し、かつ最適な合成単位の組み合わせを求めるために、ｍｉｎ－ｍａｘ ＤＰ法を基本とした合成単位波形の選択を行っている。

【００２２】ここで、ｍｉｎ－ｍａｘ ＤＰ法について簡単に説明する。

【００２３】選択された合成単位をＦ＝ｃ(１)ｃ(２)･･･ｃ(ｋ)とし、それらの合成単位を選択することにより、合成単位の各接続部分で生じる歪みをそれぞれｄ(ｃ(１))，ｄ(ｃ(２))，････，ｄ(ｃ(ｋ))とする。

【００２４】ここで、合成単位Ｆによる選択歪みをＤ(Ｆ)＝ｍａｘ［ｄ(ｃ(１))，ｄ(ｃ(２))，････，ｄ(ｃ(ｋ))］と定義する。この選択歪みＤ(Ｆ)が最小となる最適な組合せを動的計画法(ＤＰ法)を用いて求める手法をｍｉｎ－ｍａｘ ＤＰ法という。

【００２５】以下に、合成単位波形選択部（１２０）の処理の流れを説明する。

【００２６】波形選択に起因する合成音の歪みは、２つに分類することができる。一方は、合成しようとする音声波形とその音声波形に対応して選択された合成単位波形との発話環境の非適合により生ずる歪みであり、他方は合成単位の波形接続により生ずる歪みである。

【００２７】本発明では、発話環境の非適合により生ずる歪みは、音素中心付近での基本周波数D_{F0}とパワーD_{pow}、音韻継続時間長D_{dur}、文中の位置(語頭、語中、語尾)D_{prsl}の違いを数値化して評価する。

【００２８】一方、波形接続により生ずる歪みは、接続する２つの合成単位の接続部分での基本周波数差D^{c}_{F0}、パワー差D^{c}_{pow}、ケプストラムの差D^{c}_{cep}および発話環境を考慮して決定された接続の行い易さ（接続優先順位）を示す歪みD^{c}_{ph}を数値化して評価する。歪みD^{c}_{ph}は、パワーが小さく、聴感的に接続歪みが知覚されにくい接続部分ほど小さな値が設定されており、反対にパワーの大きい接続部分やスペクトルの変化の大きい接続部分等の他の合成単位との接続が行われることが望ましくない接続部分では大きな値が設定されている。

【００２９】以下に、図１に示すステップ１２０ｃで用いる歪みの評価式を数１のように定義する。

【００３０】

【数１】

$$D(F) = \sum_{i=1}^{k}(D_{F0}(c(i)) + D_{pow}(c(i)) + D_{dur}(c(i)) + D_{posi}(c(i))) + \max_{i=1\ldots k}(D^{c}_{F0}(c(i)) + D^{c}_{pow}(c(i)) + D^{c}_{cep}(c(i)) + D^{c}_{ph}(c(i)))$$

【００３１】歪みＤ(Ｆ)が最小となる合成単位の組み合わせが、最適な選択結果となる。計算時間の関係から、実際には、音韻連鎖長は最大５音素までとする。また、発話環境の非適合による歪みの評価値のみを用いて予備選択を行い、その結果選択された音素列に対して数１を計算して最適解を求める。

【００３２】ここで、合成単位波形選択部（１２０）による処理の流れを具体例を用いて説明する。

【００３３】発話文章は「回りの人も立ち上がった。」であり、特に「回り（ｍａｗａｒｉ）」について説明する。尚、本発明の実施の形態では、パワー、及びピッチから波形の合成による歪みを計算する。

【００３４】図１のステップ１２０ａで音声波形データ格納部から抽出された全ての合成単位の組み合わせの候補について、ステップ１２０ｂにおいて、全ての合成単位の組み合わせについて評価値を算出する。

【００３５】各歪みは、差の２乗により計算した。また、実際には、パワーとピッチのように異なる種類（次＊元）の歪みを加え合せる場合は、重み係数を掛けて加え合せるが、本発明の実施の形態では単に加算する。図２中のパワーおよびピッチの目標値とは、韻律パターン生成部１１で計算された目標値である。

【００３６】歪みは、２種類に分けることができる。１種類は、２つの合成単位を接続することにより生ずる歪み（図２の第１の歪み）であり、もう１種類は、目標値と波形辞書より選択した合成単位の値との差による歪み（図２の第２の歪み）である。

【００３７】第１の歪みは全ての接続部分の最大値より求め、第２の歪みは合成単位の平均値より求め、合成単位の組合せの評価値は、それら両歪みの合計とする。

【００３８】例えば、図２中の第１の歪みについては、５個所の接続部分について、数２に従って夫々歪みを計算し、その最大値３７００が第１の歪みとなる。

【００３９】

【数２】

$$\text{第1の歪}=\max\{(330-330)^2+(220-220)^2,\ (370-350)^2+(220-250)^2,\ (320-320)^2+(210-210)^2,\ (320-330)^2+(210-270)^2,\ (370-330)^2+(280-280)^2\}$$
$$=\max\{0, 1300, 0, 3700, 1600\}=3700$$

【００４０】また、図２中の第２の歪みについては、合成単位毎について、数３に従って歪みの平均を夫々計算し、その平均値３８５０を第２の歪みとなる。

※【００４１】

【数３】

※

$$\text{第2の歪}=\{(300-300)^2+(370-370)^2+(350-330)^2+(450-330)^2+(350-350)^2+(400-350)^2+(220-220)^2+(220-220)^2+(250-220)^2+(280-210)^2+(290-290)^2+(280-280)^2\}/6$$
$$=23100/6=3850$$

$$\text{評価値}=3700+3850=7550$$

【００４２】従って、この組合せの評価値は、第１の歪みの値３７００、及び第２の歪みの値３８５０の合計の７５５０となる。

【００４３】ここで、上述では、第１の歪みの値と第２の歪みの値の和によって評価値を求めたが、これ以外に以下のような２つの手法に従って、合成単位の接続における評価を行うことができる。

＜第１手法＞上述の評価値の計算を行い、最小となった評価値から一定の範囲以内（例えば、最小評価値の１．１倍迄）に属する全ての組合せについて、再度、新たな計算方法によって評価値を求め、それらの評価値の中で最小となる組合せの探索を行う。ここで、新たな計算法とは、図２の第１の歪みの値の中で、第２番目に大きな歪みを用いて計算する方法である。例えば、図２の例では、第１の歪みの中で第２番目に大きな値は１６００となり、第２の歪みが３８５０であるから、組合せの評価値は５４５０となる。

【００４４】＜第２手法＞第１の歪みについて、歪みの大きい順に重み係数を掛けた値の総和を用いる。重み係数を｛０．６，０．３，０．１｝の３つとすると、第１の歪みは、３７００×０．６＋１６００×０．３＋１３００×０．１＝２８３０となり、また第２の歪みが３８５０であるから、組合せの評価値は６６８０となる。

【００４５】ここで、第２手法では、３個の第１の歪み

が算出されたため、重み係数も３個設定したが、これには限られず重み係数の個数は、第１の歪みの個数に応じて、適宜設定すれば良い。

【００４６】また、重み係数の値は、それらの合計が１となるように設定することが好ましい。

【００４７】ところで、本発明の有効性を調べるために、本発明を用いて合成した音声と、通常のＤＰ法を用いて合成した音声との比較を行った。

【００４８】図３は、通常のＤＰ法で問題となった接続歪みの集中がｍｉｎ－ｍａｘ ＤＰ法で解消される一例である。発話文章は、「回りの人も立ち上った。」である。図３では「まわりのひ」までを表示している。

【００４９】本発明で合成した結果を図３（ａ）に、また通常のＤＰ法で合成した結果を図３（ｂ）に示す。図３（ａ）、（ｂ）における上段は音声波形を、また下段は前節で示した選択歪みを示す。選択歪みは、接続部分では接続歪みを、それ以外の位置では選択した単位と発話環境の違いによる非適合歪みを表している。音声波形の図に描かれた縦線は、接続の行なわれた個所を示している。

【００５０】通常のＤＰ法では選択歪みの総和を最小とする組合せが求められるため、長い音素列が選択された場合に接続部分での歪みが非常によく知覚されることがある。

【００５１】本実験では下図に示すように、／ａｗａｒｉ／、ｉｎｏｈ／の2つの長い音素列が選択され、接続部分の音素／ｉ／に歪みが集中していることがわかる。＊

＊【００５２】それに対して本発明では、接続歪みの最大値に注目し、その値が最も小さくなる組合せが求められるため、接続箇所は５箇所と増加し、誤差の総和も増加しているが、接続歪みの最大値は減少し、全体に歪みが分散されている。このことは、聴覚的に顕著な歪みが減少することを示している。

【００５３】

【発明の効果】以上の説明から明らかなように、本発明によれば、最適な合成単位の組合せを選択する際に、合成単位の接続部分の歪みの合計ではなく、その歪みの最大値を用い、その歪みの最大値が小さくなる合成単位の組合せを選択することにより、或る個所での接続部分への歪みの集中が緩和され、聴感的に歪みの少ない高品質の音声合成ができる効果を有する。

【図面の簡単な説明】

【図１】本発明の音声合成装置の合成単位波形選択部（１２０）の処理の流れを示した図である。

【図２】本発明の音声合成装置における合成単位の接続の組合せの評価値の算出法を示す図である。

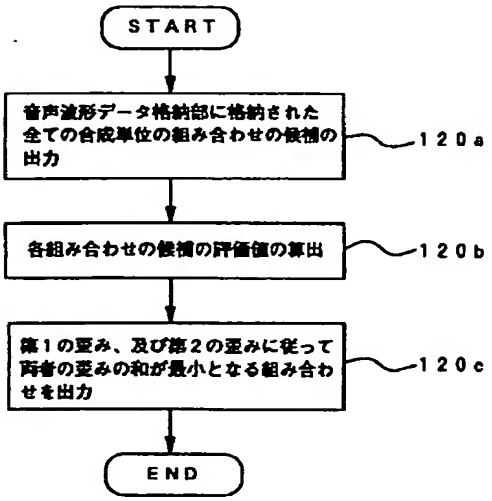
【図３】本発明、並びに従来の音声合成装置によって合成単位を接続した場合の合成結果を示す図である。

【図４】従来の音声合成装置の概略構成図を示す。

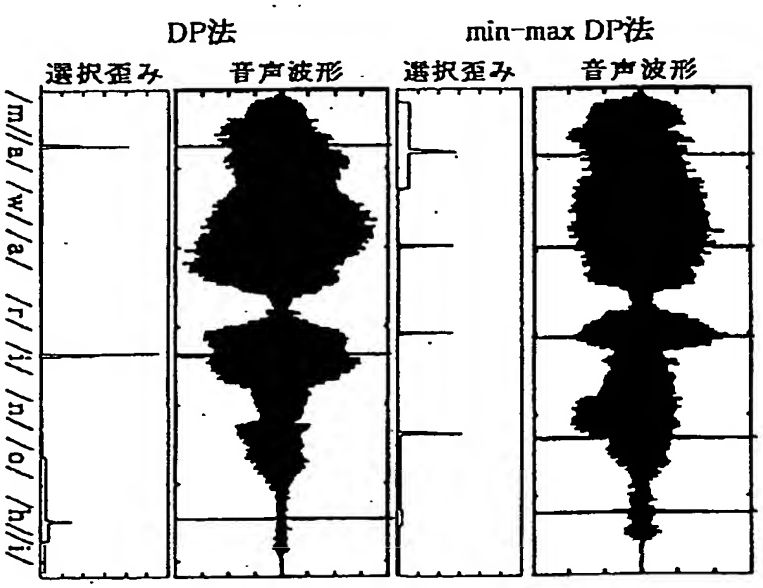
【符号の説明】

１０･･･言語処理部
１１･･･韻律パターン生成部
１２･･･合成単位波形選択部
１３･･･音声波形生成部

【図１】



【図３】



【図２】

パワー目標値		300			370			350			450			350			400	
ピッチ目標値		220			220			250			280			290			280	
		/m/			/a/			/e/			/a/			/t/			/i/	
音素片のパワー	0	300	330	330	370	370	350	330	320	320	330	320	330	350	370	330	350	300
音素片のピッチ	220	220	220	220	220	220	250	220	210	210	210	210	270	290	280	280	280	250

【図４】